1

Diagnostische Tests in der Veterinärmedizin

In diesem Kapitel werden zunächst allgemeine Grundprinzipien diagnostischer Tests ausgeführt, wobei auf Besonderheiten der serodiagnostischen Verfahren hingewiesen wird (Abschnitt 1.1). Die Definition eines Untersuchungsverfahrens als *Diagnosetest* ergibt sich nicht zuletzt durch seine Einbeziehung in einen diagnostischen Prozess, der im Folgenden grob skizziert wird (Abschnitt 1.1.1). Auf die in der Veterinärmedizin gebräuchlichen Aggregat- (Abschnitt 1.1.2) und Pooltests (Abschnitt 1.1.3), sowie auf multiple Diagnosetests (Abschnitt 1.1.4) wird gesondert eingegangen. Zum Zweck der Illustration werden Anwendungsfelder diagnostischer Tests aus den Bereichen der klinischen Veterinärmedizin (Abschnitt 1.2.1), Tierseuchenkontrolle (Abschnitt 1.2.2), Prävalenzschätzung (Abschnitt 1.2.3), epidemiologischen Untersuchung von Risikofaktoren (Abschnitt 1.2.4) und der quantitativen Risikoanalyse (Abschnitt 1.2.5) an Hand von Beispielen erläutert.

1.1 Grundprinzipien diagnostischer Tests

Ein diagnostischer Test ist ein klinisches (Adspektion, Palpation, Perkussion, Auskultation oder Mensuration) oder labormedizinisches Verfahren zum Nachweis oder zur Quantifizierung eines klinischen Symptoms, einer körpereigenen oder -fremden Substanz oder einer Gewebeveränderung mit dem Ziel der Klassifizierung des Untersuchungsobjekts in "Test positiv" ($T+$) oder "Test negativ" ($T-$). Untersuchungsobjekt in der veterinärmedizinischen Diagnostik ist das Einzeltier oder ein räumliches Tieraggregat (Bestand, Herde, etc.), in der Labordiagnostik eine oder mehrere Gewebe-, Flüssigkeits- oder Ausscheidungsproben eines Individuums oder eines Tieraggregats (Sammelprobe, *Pooltests*). Im Mittelpunkt der vorliegenden Arbeit stehen serologische Labormethoden, da diese von besonderer Relevanz für die Tierseuchendiagnostik sind. Die Formulierung "*ein* diagnostischer Test" bezieht sich immer auf eine konkrete Festlegung auf *ein* Untersuchungsverfahren (*ein* Analyt im Fall eines labordiagnostischen Tests), welches zur Diagnose *einer*

Merkmalsausprägung bei *einer* Tierart angewendet wird. Demnach würde es sich um zwei verschiedene "Tests" handeln, wenn beispielsweise ein ELISA-Verfahren zum Nachweis von *Brucella*-Antikörpern im Blutserum und in der Milch eingesetzt wird.

Die Anwendung eines Diagnosetests ist dann sinnvoll, wenn der *wahre* Status des Untersuchungsobjekts, "Status positiv" ($S+$) oder "Status negativ" ($S-$), unbekannt ist, jedoch mit einer akzeptablen Fehlerrate durch die diagnostische Klassifikation beurteilt werden kann. Die Abschätzung verschiedener Fehlerraten bei der diagnostischen Klassifikation ist das Ziel der *Evaluierung* und wird in Abschnitt 3.3 erläutert. Im Folgenden werden einige Aspekte der oben genannten Definition näher betrachtet.

Der zu diagnostizierende Status S betrifft häufig das Vorhandensein einer Krankheit oder einer Infektion. Jedoch kann auch die Bestimmung von unveränderlichen, angeborenen (z.B. Geschlechtsdiagnostik bei Ziervögeln), veränderlichen physiologischen (z.B. Trächtigkeitsdiagnostik) oder immunologischen (z.B. Nachweis von protektiven Antikörpern nach Impfung) dichotomen[1] Merkmalen Ziel der diagnostischen Testung sein.

Dem Konzept der dichotomen Ausprägung des wahren Status S im Sinn einer Diagnosestellung steht nicht immer eine "natürliche" Dichotomie der zu Grunde liegenden klinischen oder biologischen Phänomene gegenüber. Eine Infektionskrankheit beispielsweise kann, abhängig vom zeitlichen Stadium (Inkubation, Generalisation, Manifestation bei einer zyklischen Infektionskrankheit) und der immunologischen Abwehrlage des Tiers, in unterschiedlicher klinischer Ausprägung vorliegen (z.B. perakut, akut, chronisch, latent), wobei fließende Übergänge zwischen den Ausprägungsformen möglich sind. Der wahre *positive* Status eines Tiers wäre demnach nur durch eine Variable mit kategorialem oder kontinuierlichem Skalenniveau präzise beschreibbar. Auch die *negative* Ausprägung des wahren Status kann häufig weiter differenziert werden. In der Infektionsdiagnostik ist hier ein etwaiger Impfschutz, eine überstandene Infektionsepisode mit dem homologen Erreger oder das Vorliegen heterologer Infektionen zu berücksichtigen. Der zu diagnostizierende dichotome Status S ist daher als eine nach den Erfordernissen der Fragestellung gewählte, "künstliche" Unterteilung der Population zu verstehen, die jedoch bei der Evaluierung und Anwendung von Tests übereinstimmend vorgenommen werden muss.

Auch die Ergebnisse des diagnostischen Tests liegen häufig nicht primär in dichotomer Form vor. Viele in der veterinärmedizinischen Infektionsdiagnostik angewendete Tests (Abschnitt 1.2) liefern zunächst ordinal-skalierte[2] oder

[1] Merkmale mit lediglich zwei möglichen, sich gegenseitig ausschließenden und (abgesehen von der Möglichkeit fehlender Werte) gemeinsam erschöpfenden Ausprägungsformen werden als *dichotom* bezeichnet.

[2] Merkmale, deren Ausprägungen in einer natürlichen Rangfolge angeordnet werden können, werden als *ordinal-skaliert* oder *polytom* (auch polychotom) bezeichnet (z.B. Titerstufen).

quantitative[3] Ergebnisse, die an Hand eines Grenzwerts (*Cut-off*, Abschnitt 5.1) dichotomisiert werden.

Die doppelt dichotome Darstellung des wahren Status und des Ergebnisses eines Diagnosetests ist ein Charakteristikum aller diagnostischen Tests, beruht jedoch auf einer kontextspezifischen Definition der Ausprägungen von S und T. Dies sei am Beispiel eines diagnostischen Antikörper-ELISAs verdeutlicht (Abb. 1.1). Die biochemischen Vorgänge beim Nachweis von Immunkomplexen im ELISA sind wohlbekannt. Die bei der Untersuchung einer Probe wirksamen Stör- und Einflussfaktoren (Abschnitt 2.2.3) sind dagegen in der Regel nicht vollständig bekannt, so dass die Darstellung der Analysemethode als "*Black box*" gerechtfertigt erscheint (Abb. 1.1).

Abhängig vom Anwendungskontext ist entweder die Diagnose einer Krankheit oder einer Infektion von Bedeutung. Darüber hinaus kann die diagnostische Abgrenzung von immunen, vakzinierten oder von anderen Erkrankungen oder Infektionen betroffenen Tieren sowie eine Erkennung von klinisch inapparent (z.B. latent) infizierten Tieren relevant sein. Die Definition von $S+$ und $S-$ ist also vom diagnostischen Kontext abhängig. Die Auswahl eines diagnostischen Markers (hier Konzentration spezifischer Antikörper im Blutserum) erfolgt aufgrund wissenschaftlich plausibler Erwägungen[4]. Da der Marker nicht direkt zu beobachten ist, müssen Messverfahren (hier ELISA) eingesetzt werden, die mit systematischen und zufälligen Fehlern behaftet sind, die an dieser Stelle nicht weiter erläutert werden sollen. Als Ergebnis der Messung wird ein Extinktionswert (oder optische Dichte) ermittelt, der an Hand mitgeführter Kontrollproben "*normalisiert*"[5] und durch Anwendung eines Grenzwerts dichotomisiert wird. Hierdurch wird der Status $T+$ oder $T-$ des Untersuchungsobjekts definiert.

An dieser Stelle wird deutlich, dass in der Regel eine fehlerfreie Klassifizierung durch den Test nicht zu erwarten ist. Ein Nachweis von Antikörpern ($T+$) bei gesunden, nicht infizierten Tieren mit künstlich oder natürlich erworbener humoraler Immunität ($S-$) oder aber ein negatives Testergebnis ($T-$) bei Tieren im perakuten oder akuten Infektionsstadium ($S+$) wäre im *diagnostischen* Sinn als *falsch* anzusehen, obwohl ein Messfehler im *analytischen* Sinn nicht vorliegt. Der Grad der Übereinstimmung zwischen der *wahren* und der *diagnostischen* Klassifikation bei der Untersuchung von n Testobjekten kann an Hand der Besetzungshäufigkeiten einer Befundmatrix (Vierfeldertafel im Fall der Dichotomie von S und T) quantifiziert werden (Abb. 1.2). Hierbei ist es üblich, die Häufigkeiten von beobachteten richtig positiven, falsch positiven,

[3] Merkmale, denen eine kontinuierliche Intervall- oder Verhältnisskala zu Grunde liegt, werden als *quantitativ* bezeichnet (z.B. ELISA-Extinktionswerte).

[4] Man kann an dieser Stelle von einer zu fordernden *Konstruktvalidität* sprechen.

[5] Eine Korrektur von systematischen Messfehlern durch mitgeführte Kontrollproben wird von Wright et al. (1993 [258]) und Wright und Zhou (1999 [256]) beschrieben und dort als "*Normalisierung*" bezeichnet. Hierzu ist kritisch anzumerken, dass auf diese Weise Zufallsschwankungen des Kontrollwerts auf die "korrigierten" Proben-Messwerte übertragen werden.

(1) Krankheitsstatus

nicht erkrankt	erkrankt

(2) Infektionsstatus

nicht infiziert		infiziert	
naiv	andere Erkrankungen	subklinisch	perakut
vakziniert	andere Infektionen	latent	akut
immun		okkult, toleriert	chronisch

(3) Serologischer Marker

Konzentration von spezifischen Antikörpern im peripheren Blut

(4) Antikörper-ELISA

Black box

(5) ELISA Messwert

Optische Dichte oder Indexwert bezogen auf Positivkontrolle

(6) Serologische Diagnose

ELISA negativ	ELISA positiv

Abb. 1.1. Schematische Modellvorstellung über die Plausibilität eines Antikörper-ELISAs zur Diagnostik einer Infektionskrankheit. Die Krankheitsausprägung (1) wird ursächlich auf einen Infektionszustand zurückgeführt (2), für den die Existenz eines patho-physiologischen Markers (Infektionsantikörper) (3) postuliert wird. Die immunologische Analyse des Markers erfolgt in einem Prüfsystem (ELISA) (4), dessen analytische Leistungsfähigkeit und Robustheit nicht vollständig bekannt sind ("*Black box*"). Die erhobenen Messwerte, nach Korrektur für bestimmte Messfehler (5), werden an Hand eines geeigneten Grenzwerts diagnostisch beurteilt (6).

falsch negativen und richtig negativen Diagnosen mit den Buchstaben a, b, c und d zu bezeichnen. Die Konstruktion solcher Befundmatrizen nimmt eine zentrale Rolle bei der Evaluierung eines Diagnosetests ein (Abschnitt 3.3).

1.1.1 Diagnostischer Prozess

Die Durchführung eines diagnostischen Tests erfolgt nach festgelegten Regeln, die im Sinn eines modernen Qualitätsmanagements in einer Standardarbeitsanweisung (*Standard operating procedure*, SOP) dokumentiert werden. Die SOP für einen diagnostischen Test enthält verbindliche Vorschriften

	Wahrer Status	
	$S+$	$S-$
Testergebnis $T+$	a	b
$T-$	c	d

Abb. 1.2. Die Befundmatrix gibt die absoluten Häufigkeiten richtig positiver (a), falsch positiver (b), falsch negativer (c) und richtig negativer (d) Diagnosen eines Tests wieder, basierend auf der Ausprägung des wahren Status (S) und des Tests (T). Die Tafelsumme sei n.

zur Ausführung von präanalytischen (z.B. Vorbehandlung der Proben), analytischen und postanalytischen Arbeitsschritten. Zu den postanalytischen Arbeitsschritten der Testdurchführung gehören die Befunddokumentation sowie die Mitteilung des Befunds an die auftraggebende Stelle. Der Prozess der medizinischen Diagnosestellung ist durch folgende Elemente charakterisiert:

(i) Formulierung einer Untersuchungshypothese;
(ii) Entscheidung zur Durchführung eines Diagnosetests;
(iii) Entscheidung zur Auswahl eines oder mehrerer Diagnosetests;
(iv) Entscheidung zur Auswahl von Testobjekten;
(v) Entscheidung zur Auswahl des Untersuchungsmaterials;
(vi) Durchführung des Diagnosetests einschließlich vorbereitender Schritte;
(vii) Beurteilung des Testbefunds.

Am Anfang des diagnostischen Prozesses steht die Formulierung einer Untersuchungshypothese (i), die auf mehr oder weniger umfangreichen Vorinformationen beruht. Zu diesen können die Ergebnisse von bereits durchgeführten Tests zählen, welche auch die Entscheidungen (ii) und (iii) maßgeblich beeinflussen. In der statistisch orientierten Literatur wird die auf Vorinformationen beruhende Wahrscheinlichkeit für das Vorliegen des fraglichen Status als *Prioriwahrscheinlichkeit* bezeichnet. Lestin (1995 [148], S. 3) weist in diesem Zusammenhang auf die Bedeutung von Ausschlussdiagnosen hin. Die Auswahl einer von vielen möglichen Untersuchungshypothesen hat bereits einen "Entscheidungscharakter". Wie an anderer Stelle (Abschnitte 1.2.1 und 6.1) erläutert werden wird, sind Vorinformationen, insbesondere in der Form von quantitativen Abschätzungen, für die Entscheidung (vi) von zentraler Bedeutung.

Die Entscheidung (ii) erfolgt aufgrund einer Kosten-Nutzen-Abwägung, bei der die finanziellen Aufwendungen für die Testung sowie Risiken durch invasive Testverfahren dem erwarteten Nutzen gegenübergestellt werden. Zu den Risiken von diagnostischen Tests zählen auch die Konsequenzen von fehlerhaften (falsch positiven und falsch negativen) Testergebnissen. Die formalisierte

Betrachtungsweise von Schritt (ii) ist eine klassische Anwendung der Entscheidungsanalyse, bei der alle Entscheidungsalternativen unter Berücksichtigung von Kosten und Nutzen systematisch erfasst und numerisch ausgewertet werden (siehe hierzu Petitti, 1994 [199]). Auch im Bereich der veterinärmedizinischen Diagnostik hat die Entscheidungsanalyse Anwendung gefunden (Smith, 1993 [230]; Collins et al., 1996 [47], Greiner und Baumann, 1998 [87]; Smith und Slenning, 2000 [232]; Abschnitt 4.2.3; Anhang B.5). Die Entscheidung, einen Test durchzuführen, ist abhängig von der Prioriwahrscheinlichkeit sowie vom Grad der Sicherheit, mit der letztere angegeben werden kann. In diesem Zusammenhang wurde in der medizinischen Diagnostik vor einer unkritischen Anforderung von Tests gewarnt. Falsch positive Untersuchungsergebnisse von nicht indizierten Tests erfordern häufig eine weitergehende diagnostische Abklärung, die unwirtschaftlich und im medizinischen Kontext risikobehaftet sein kann (*Diagnostic cascade effect*, Mold und Stein, 1986 [176]).

Die Auswahl eines oder mehrerer Tests in Schritt (iii) ist durch den aktuellen wissenschaftlichen Kenntnisstand der klinischen Propädeutik und der zur Diagnostik beitragenden nicht-klinischen Fächer vorgegeben. Durch die Auswahl eines labormedizinischen Tests ist eine Festlegung auf eine *Prüfgröße*, d.h. eine Messvariable des Analyten in der Probe, gegeben. Im Zusammenhang mit der Tierseuchendiagnostik sind einschlägige Rechtsvorschriften zu beachten. International verbindliche Standards für die Auswahl von diagnostischen Tests in der Tierseuchendiagnostik sind im "*Manual of Standards for Diagnostic Tests and Vaccines*" (OIE, 2000 [193]) niedergelegt. Die grundsätzliche Eignung eines Tests, seine diagnostische Qualität für die spezifische Anwendungssituation sowie die direkten und indirekten Kosten und Risiken der Anwendung sind wichtige Auswahlkriterien. Bei einer zusammenfassenden Erwägung der Entscheidungen unter (ii) und (iii) sollte der mögliche Informationsgewinn durch den Diagnosetest gegenüber der bereits vorliegenden a priori Information berücksichtigt werden. Lässt die Vorinformation unter (i) bereits eine sichere Diagnose zu, so ist die Anwendung weiterer Tests obsolet[6]. Wenn Ergebnisse bereits durchgeführter Tests Bestandteil der Vorinformation sind, so ist ein Informationsgewinn durch Anwendung eines weiteren Tests nur dann zu erwarten, wenn der zweite Test eine höhere diagnostische Zuverlässigkeit hat oder bedingt unabhängig vom ersten Test ist (Politser, 1982 [202]; Chiecchio et al., 1994 [37]; Gardner et al., 2000 [76]; Abschnitt 4.2.4).

Die Entscheidung (iv) zur Auswahl von Testobjekten spielt insbesondere bei der Diagnostizierung von Tieraggregaten (z.B. Herde, Käfigbesatz, Fischtank) und bei epidemiologischen Fragestellungen eine bedeutende Rolle und ist eng mit dem Komplex der Stichprobengewinnung verknüpft. Die Stichprobenziehung, zu der Cannon und Roe (1982 [31]) detaillierte Hinweise geben, ist jedoch von den Untersuchungszielen abhängig. So ist eine

[6] Dies gilt nicht für die Anwendung von Tests zu Evaluierungszwecken (Abschnitt 3.3).

zufallsbasierte (zumindest jedoch eine systematische) Auswahl von Testobjekten erforderlich, um zu einer unverzerrten Schätzung der Prävalenz im Tieraggregat zu gelangen. Untersuchungen zum Nachweis des "Freiseins" eines Tierbestands von einer Tierseuche oder auch die aktive Suche nach Krankheitsfällen im Rahmen einer Fall-Kontroll-Studie dagegen haben zum Ziel, mit einer möglichst großen Wahrscheinlichkeit ein Tieraggregat korrekt zu klassifizieren oder Krankheitsfälle mit einer hohen Nachweisrate aufzudecken. Christensen und Gardner (2000 [40]) weisen darauf hin, dass bei aufdeckenden Untersuchungen Abweichungen von einer zufallsbasierten (oder systematischen) Auswahl motiviert sein können, um vornehmlich diejenigen Individuen zu testen, die aufgrund von Risikofaktoren oder klinischen Symptomen das Vorhandensein des fraglichen Status erwarten lassen. Generell wird die Anzahl der zu untersuchenden Tiere von der gewünschten statistischen Sicherheit, der Prävalenz, den Testeigenschaften und den Testkosten abhängen.

Die Auswahl eines geeigneten Probenmaterials ist für Tests relevant, die auf klinisch-chemischen, serologischen oder molekularbiologischen Methoden beruhen. Die Entscheidung (v) trägt daher den technischen Erfordernissen des angewendeten Diagnosetests Rechnung. Das Probenmaterial muss die nachzuweisende Substanz (*Analyt*) in einer Konzentration enthalten, die über der analytischen Nachweisgrenze liegt. Störeinflüsse durch nicht analysierte Komponenten der Probe (*Probenmatrix*) sollten weitgehend ausgeschlossen sein. Bei der Auswahl des Probenmaterials werden darüber hinaus möglichst standardisierte, kostengünstige und wenig invasive Verfahren angestrebt. Ähnliche Entscheidungskriterien gelten für Tests, die nicht auf analytischen Verfahren beruhen, wie beispielsweise mikroskopische Direktnachweisverfahren von Infektionserregern.

Bei einem labormedizinischen Diagnosetest beinhaltet Schritt (vi) eine präanalytische (evtl. Vorbehandlung des Tiers, Probenentnahme, Lagerung, Versand und Vorbereitung der Probe) und analytische Arbeitsphase (Einbringen der vorbehandelten Probe in die Analyse, Durchführung und Ablesen des Tests nach SOP). Bei einem Diagnoseverfahren, welches direkt am Tier durchgeführt werden kann (*Pen-side* Diagnostik), entfallen Aufwendungen für Lagerung und Transport von Proben. Allerdings sind die Anforderungen an die Robustheit (Abschnitt 2.1) solcher Verfahren besonders hoch.

Der abschließende und wichtigste Schritt (vii) des diagnostischen Prozesses ist die diagnostische Beurteilung der Testbefunde unter Berücksichtigung der Prioriwahrscheinlichkeit und der diagnostischen Güte des Tests. Der diagnostische Befund ist eine der Grundlagen für das medizinische Handeln, welches durch die Kontextbedingungen (medizinische, wirtschaftliche, ethische und soziale Konsequenzen des Handelns) und durch den Grad der (Un)Sicherheit über die Richtigkeit des Testbefunds und über die Konsequenzen des Handelns geprägt ist.

Der hier skizzierte reguläre Prozess der diagnostischen Testung gilt sowohl für die klinische Diagnosefindung als auch – im übertragenden Sinn – für nicht-klinische Anwendungssituationen.

1.1.2 Aggregattests (Herdentests)

Zur Bekämpfung von Tierseuchen und Zoonosen sowie im Rahmen der Qualitätssicherung in der Tierproduktion werden Zertifizierungen von Betrieben über das Freisein von bestimmten Infektionen durchgeführt (Abschnitt 1.2.2). In diesen Fällen ist jedoch nicht das Einzeltier, sondern der gesamte Tierbestand eines Betriebs (*Tieraggregat*) als diagnostische Einheit aufzufassen.

Als Aggregattests (*Aggregate testing*) werden solche diagnostischen Tests bezeichnet, die an Einzeltieren ausgeführt werden, jedoch auf die diagnostische Klassifizierung des Tieraggregats abzielen (Christensen and Gardner, 2000 [40]). Jegliche Form einer haltungs- oder produktionstechnischen oder sonstigen Zusammenstellung von Tieren, wie beispielsweise eine Herde, ein Bestand an Legehennen, ein Fischtank, ein Bienenvolk oder eine Tiergruppe in einer Quarantänestation kann als Aggregat aufgefasst werden. In der Literatur hat sich jedoch der Begriff "Herdentest" als allgemeine Bezeichnung für die Testung jedweder Aggregate durchgesetzt. Dieser Vereinfachung folgend, wird von nun an der Begriff der "Herde" in diesem Zusammenhang ohne Bezugnahme auf eine bestimmte Tierart oder Haltungsform verwendet. Für die positive und negative Klassifikation einer Herde mit einem Herdentest (T_H) werden im Folgenden die Bezeichnungen T_H+ und T_H- verwendet. Die diagnostische Klassifizierung erfordert eine Entscheidungsregel, bei welcher Anzahl (oder welchem Anteil) von positiven Reagenten das Aggregat den Status T_H+ erhält. Im einzelnen bezeichne

N : die Gesamtanzahl der Tiere in der Herde;

n : die Anzahl der untersuchten Tiere in der Herde;

c : die Mindestanzahl positiver Individualtestergebnisse für T_H+;

Y : die tatsächliche Anzahl positiver Individualtestergebnisse.

Der Herdentest liefert zunächst eine quantitative (binomial-skalierte) Messgröße Y/n mit $n+1$ möglichen, diskreten Beobachtungen $\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}$. Folglich wird ein Grenzwert, wie beispielsweise $\frac{c}{n}$, benötigt, um zu einer diagnostischen Entscheidung zu gelangen. Die Diagnose T_H+ wird genau dann gestellt, wenn $Y \geq c$ ist. Der Status T_H+, bzw. T_H- gilt für alle Tiere einer Herde mit dem jeweiligen Resultat des Herdentests. Der diagnostische Status von Einzeltieren ist hierbei nicht von Interesse. Abbildung 1.3 zeigt eine schematische Darstellung eines Herdentests. Die Evaluierung von Aggregattests wird unter Abschnitt 4.1 erläutert.

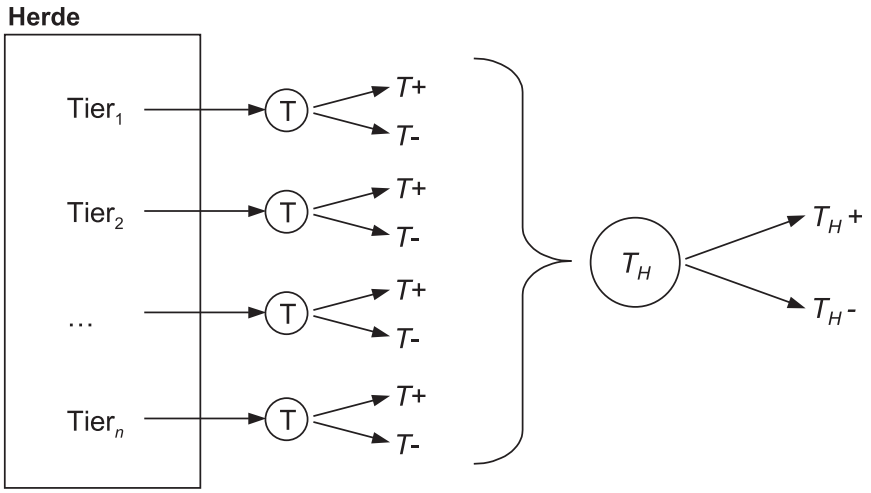

Abb. 1.3. Schematische Darstellung eines Herdentests (T_H). Der diagnostische Test (T) wird an n Einzeltieren der Herde durchgeführt und die aggregierten Ergebnisse zur Klassifizierung der Herde an Hand einer Entscheidungsregel (z.B. T_H+ wenn mindestens 1 Tier ein positives Testresultat $T+$ aufweist) verwendet.

1.1.3 Pooltests ("Bulktests")

Gepoolte Tests basieren auf der stofflichen Vermengung von Untersuchungsmaterial einzelner Tiere vor der Testung und sind daher auf labordiagnostische Verfahren beschränkt. Sie sind potenziell bei Nachweisverfahren mit hoher analytischer Sensitivität einsetzbar und führen hier zu einer drastischen Einsparung von Testkosten. Es sollte jedoch der Verlust an analytischer Sensitivität durch den eintretenden Verdünnungseffekt nicht vernachlässigt werden.

Im Allgemeinen ist jedoch nicht die Sammelprobe, sondern das Einzeltier, die Herde oder die Population von klinischem oder epidemiologischem Interesse. Daher werden – sofern "Rückstellproben" von Einzeltieren zur Verfügung stehen – entweder "rückwärtsgerichtet" alle Individuen mit positivem Pooltest nachkontrolliert oder aber die Ergebnisse des Pooltests im Sinn eines Herdentests zusammengefasst. In Abbildung 1.4 ist ein solcher Fall schematisch dargestellt. Es sei darauf hingewiesen, dass ein Pooling von Proben verschiedener Herden nicht mit der Definition eines Herdentests zu vereinbaren ist (Christensen und Gardner, 2000 [40]).

Die positive und negative Ausprägungskategorie eines Pooltests (T_P) wird hier mit T_P+ und T_P- bezeichnet. Analog zu der Situation bei einem Herdentest nehmen alle Individuen den jeweiligen Status des Pools an, wobei Aussagen über ihren diagnostischen Status $T+$ oder $T-$ nicht getroffen werden können.

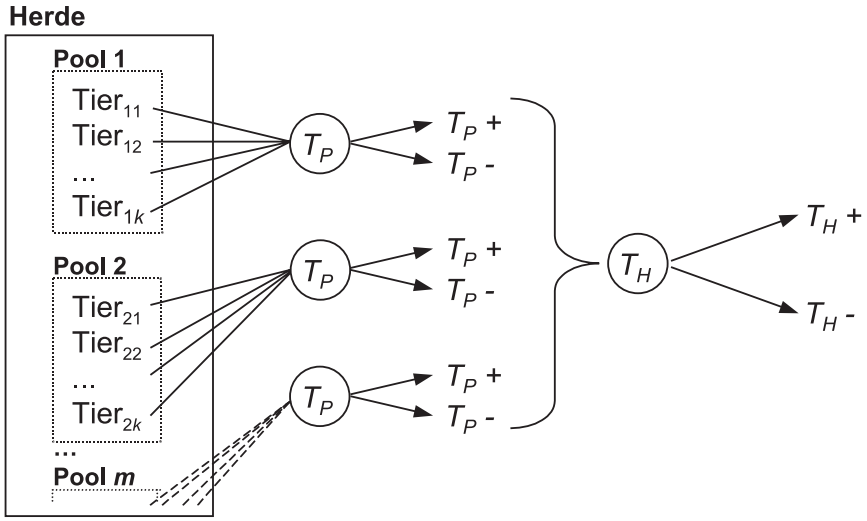

Abb. 1.4. Schematische Darstellung eines Pooltests (T_P) in Kombination mit einem Herdentest (T_H). Der Pooltest führt zur Klassifizierung (T_P+, T_P-) von m Vereinigungsproben (Pools) von jeweils k Tieren. Die Herde wird an Hand einer Entscheidungsregel (z.B. T_H+ wenn mindestens 1 Pool ein positives Testresultat T_P+ aufweist) klassifiziert.

1.1.4 Multiple Tests

Eher selten basiert eine abschließende Diagnose auf einem einzigen diagnostischen Test. Durch Verwendung einer geeigneten *Teststrategie* ist es häufig möglich, verschiedene (fehlerbehaftete) diagnostische Tests so miteinander zu verknüpfen, dass der resultierende "multiple Test" (T_M) über eine bessere diagnostische Leistungsfähigkeit verfügt als die Einzeltests. Im Folgenden bezeichne s die Anzahl der durchgeführten Tests T_i, $i = 1, \ldots, s$ und T_M+ und T_M- die Entscheidungskategorien des multiplen Tests.

Eine Möglichkeit der Anwendung multipler Tests ist die *parallele* Testdurchführung (Abb. 1.5). Hierbei werden für jede Untersuchungseinheit die Ergebnisse von T_i erhoben und an Hand einer Entscheidungsregel zusammengefasst. Die s möglichen, auf einem Grenzwertprinzip basierenden Regeln für Paralleltests, RP, sind

RP1: T_M+ wenn mindestens ein Test von s Tests positiv ist,
RP2: T_M+ wenn mindestens zwei von s Tests positiv sind,
$\vdots$

RPs: T_M+ wenn alle s Tests positiv sind,

und geben die notwendige Anzahl positiver Tests für eine T_M+ Diagnose an. Die Regel RP1 ist am weitesten verbreitet und wird von einigen Autoren

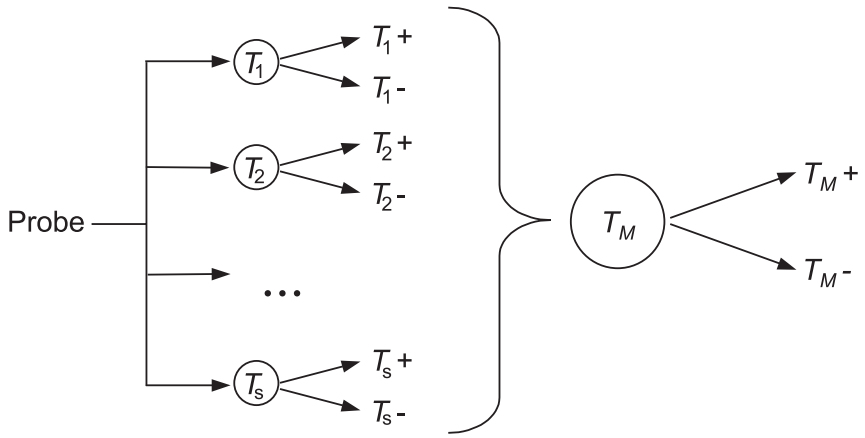

Abb. 1.5. Schematische Darstellung eines multiplen Tests (T_M) mit einer parallelen Durchführung und Interpretation von s einzelnen Diagnosetests. Die s Einzelbefunde werden zusammengefasst und nach einer festgelegten Entscheidungsregel (z.B. R1: T_M+ wenn mindestens ein Einzeltest positiv ist) bewertet.

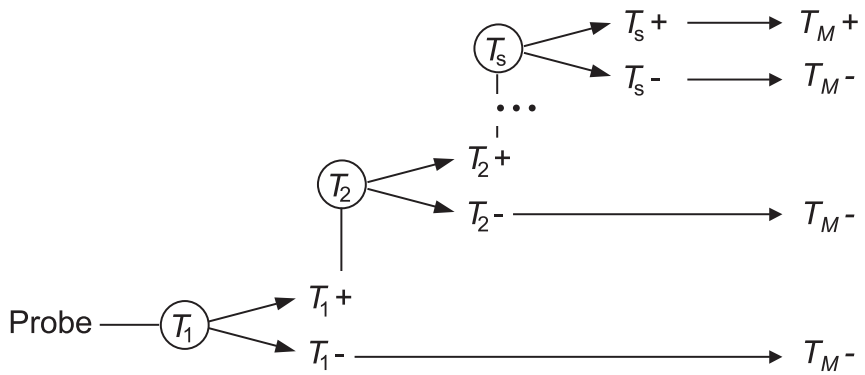

Abb. 1.6. Schematische Darstellung eines multiplen Tests (T_M) mit einer sequenziellen Durchführung und Interpretation von s einzelnen Diagnosetests. T_M wird als positiv bewertet, wenn alle s Einzeltests positiv sind. Die Testung bricht mit dem ersten negativen Einzelbefund ab und wird dann als T_M- gewertet (positiv-sequenzieller Test, $^+$S).

vereinfachend mit dem Prinzip der Paralleltestung gleichgesetzt (Gardner et al., 2000 [76]). Neben diesen Grenzwertregeln sind auch komplexere Regeln möglich. Als Beispiel sei angenommen, dass $s = 3$ ist und von T_1 falsch negative aber keine falsch positiven Resultate zu erwarten sind, während

bei T_2 und T_3 mit falsch positiven Resultaten gerechnet werden muss. Eine sinnvolle Regel könnte dann "T_M+ wenn T_1+ oder (T_2+ und T_3+)" sein.

Eine weitere Realisierungsmöglichkeit für multiple Tests ist die *sequenzielle* Testung, bei der erst nach Vorliegen des jeweiligen Testresultats entschieden wird, ob ein weiterer Test zur Bestätigung durchgeführt wird. Die Reihenfolge der Tests sowie die Abbruchkriterien (positiver oder negativer Einzelbefund, maximale Anzahl von Einzeltest, s) sind in einer Verfahrensregel festgelegt. Im Folgenden werden multiple Tests mit dem Abbruchkriterium eines negativen Einzeltestbefunds (d.h. nur positive Testbefunde werden weiter abgeklärt, Abb. 1.6) als "*positiv-sequenzielle*" ($^+$S) Diagnosetests bezeichnet. Mitunter wird die positiv-sequenzielle Testprozedur mit dem Prinzip der sequenziellen Testung gleichgesetzt (Gardner et al., 2000 [76]). Ein weit verbreitetes Beispiel für eine positiv-sequenzielle Testung ist das Screening in Kombination mit diagnostischer Abklärung von positiven Reagenten. Ist das Abbruchkriterium ein positiver Einzeltestbefund (d.h. nur negative Testbefunde werden weiter abgeklärt), liegt ein "*negativ-sequenzieller*" ($^-$S) Diagnosetest vor. Diese Teststrategie findet häufig in der Medizin bei der Abklärung unklarer Symptomatiken Anwendung. In beiden Varianten der sequenziellen Teststrategie besteht das Risiko eines Kaskaden-Effekts (Abschnitt 1.1.1). Hierunter werden direkte und indirekte Kosten der weiteren diagnostischen Abklärung verstanden, die auf Grund fehlerhafter Ergebnisse von Voruntersuchungen zu Stande kommen. Eine sinnvolle Auswahl, Reihenfolge und Begrenzung der maximalen Anzahl von Einzeltests sollte dieses Risiko minimieren. Es gibt zwei mögliche Verfahrensregeln für sequenzielle Diagnosetests mit vorgegebener Anzahl und Reihenfolge von s Einzeltests.

R$^+$Ss: T_M- sobald der erste aus einer Serie von s Tests negativ ist,
T_M+ wenn alle s Tests positiv sind.
R$^-$Ss: T_M- wenn alle s Tests negativ sind,
T_M+ sobald der erste aus einer Serie von s Tests positiv ist.

Die Abbruchkriterien der sequenziellen Diagnosetests können zu einer Einsparung von Testkosten gegenüber der Paralleltestung führen. Denkbar sind auch Kombinationen aus sequenziellen und parallelen Tests. Beispielsweise könnte T_2 eines sequenziellen Tests eine Paralleltestung sein.

Im Zusammenhang mit multiplen Tests kann unter dem Begriff der *Teststrategie* die Auswahl von Einzeltests, deren Verwendungsmodus (parallel, sequenziell, Kombination) sowie die Festlegung einer Interpretationsregel verstanden werden. Für alle solche Teststrategien gilt, dass der Informationsgewinn von der Sensitivität und Spezifität der beteiligten Tests und von der Korrelation von Testfehlern abhängt (Gardner et al., 2000 [76]). Auf die Rahmenbedingungen bei der Beurteilung von multiplen Diagnosetests sowie auf die Möglichkeiten der Evaluierung multipler Tests wird an anderer Stelle eingegangen (Abschnitt 4.2).

1.2 Anwendungsfelder serodiagnostischer Tests in der Veterinärmedizin

Die vielfältigen Anwendungsfelder für serologische Testverfahren in der Veterinärmedizin werden im Folgenden an Hand von ausgewählten Beispielen aus dem Bereich der klinischen Veterinärmedizin (Abschnitt 1.2.1), der Tierseuchenkontrolle (Abschnitt 1.2.2) und der epidemiologischen Forschung erläutert. Bei den hier berücksichtigten epidemiologischen Fragestellungen handelt es sich um die Abschätzung einer Prävalenz (Abschnitt 1.2.3) und eines Risikofaktors (Abschnitt 1.2.4). Die jeweiligen Implikationen von Testfehlern werden kurz erläutert. Die Auswahl der hier aufgeführten Anwendungsfelder ist auf solche begrenzt, in denen es vornehmlich um eine Untersuchung von Tieren, Tieraggregaten oder Populationen geht.

1.2.1 Klinische Veterinärmedizin

Für die Erkennung oder den Ausschluss vieler bakterieller, viraler oder parasitärer Infektionen in der klinischen Veterinärmedizin sind serologische Untersuchungsverfahren angezeigt und daher Bestandteil der Routinediagnostik.

Beispiel 1.1 (Diagnostik der Herzwurm-Erkrankung beim Hund)
*Smith und Slenning (2000 [232]) beschreiben den veterinärmedizinischen Entscheidungsprozess bei der klinischen Serodiagnostik (Antigen-ELISA) der Dirofilariose des Hunds, verursacht durch den Herzwurm (*Dirofilaria immitis*). Zu Grunde gelegt wurde eine typische, jedoch nicht pathognomonische klinische Symptomatik, die auch bei 14 weiteren Differentialdiagnosen auftritt. Die Faktoren des Entscheidungsprozesses (Prioriwahrscheinlichkeit bei symptomatischen Hunden, diagnostische Güte des Antigen-ELISAs, Wahrscheinlichkeit von Komplikationen bei Behandlung, Überlebenswahrscheinlichkeit in Abhängigkeit von der Infektion und Behandlung) wurden quantifiziert. Unter den getroffenen Annahmen erwies sich die Entscheidung für einen Diagnosetest als sinnvoll. Durch eine Sensitivitätsanalyse*[7] *wurde gezeigt, dass bei hoher Prioriwahrscheinlichkeit, insbesondere in Kombination mit einer hohen Wahrscheinlichkeit für falsch negative Ergebnisse des ELISAs, eine sofortige Behandlung ohne serologische Testung angezeigt ist.*

Die Serodiagnostik der caninen Filariose ist ein klassisches Beispiel für eine Kosten-Nutzen-Abwägung in der klinischen Serodiagnostik. Die formalisierte Betrachtungsweise des diagnostischen Prozesses erfordert genaue Kenntnisse

[7] Eine Sensitivitätsanalyse untersucht den Einfluss einer Modellannahme auf das Ergebnis der Modellierung.

der Epidemiologie und Klinik (Symptomatik, Therapie). Die Prioriwahrscheinlichkeit für den Herzwurmbefall wurde von den Autoren in Beispiel 1.1 mit 1/15 angegeben, unterstellt also eine exakt gleiche Prioriwahrscheinlichkeit der Dirofilariose und 14 weiterer Differentialdiagnosen bei den fraglichen Patienten. Die canine Dirofilariose tritt in bestimmten endemischen Gebieten Afrikas, Asiens, Australiens, Europas, Nord- und Südamerikas auf (Lok, 1988 [154]). Ebenso wäre die epidemiologische Situation der differenzialdiagnostischen Ätiologien zu klären. Möglicherweise würde sich die Einschätzung der Prioriwahrscheinlichkeit für Dirofilariose nach einer differenzierten Expositionsanamnese ändern. Auch aus klinischer Sicht könnte die Entscheidungsanalyse weiter differenziert werden. So ist die Symptomatik, die Aussagekraft des Antigen-ELISAs und auch die Prognose der caninen Dirofilariose abhängig vom Infektionsstadium (Schrey, 1996 [220]). Generell ist davon auszugehen, dass Diagnosefehler in der klinischen Veterinärmedizin zu Schäden durch eine nicht-optimale Behandlung des Patienten führt.

1.2.2 Kontrolle von Tierseuchen und Zoonosen

Aus wirtschaftlichen und veterinärmedizinischen Gründen sowie auch aus Gründen des gesundheitlichen Verbraucherschutzes ist es angezeigt, bestimmte Infektionen, Krankheiten oder Faktoren, die zu einer Leistungsminderung bei Nutztieren führen, auf Populationsebene quantitativ zu erfassen. Wegen der überragenden Bedeutung der Tierseuchen und Zoonosen in diesem Zusammenhang wird im Folgenden besonders auf diese Bezug genommen. Eine quantitative Erfassung der zeitlichen und räumlichen Verteilung einer Tierseuche ohne regulatorische Konsequenzen aus den Untersuchungsbefunden wird als *Monitoring* bezeichnet, während eine Überwachung (*Surveillance*) in der Regel regulatorische Maßnahmen einschließt. Wird die gesamte Population einer empfänglichen Tierart in die Untersuchung eingeschlossen, so spricht man von einem *Screening*[8] (Greiner und Gardner, 2000 [98]).

Beispiel 1.2 (Salmonellose-Bekämpfungsprogramm in Dänemark)
Ein landesweites Bekämpfungsprogramm gegen die Salmonellose bei Schweinen in Dänemark wurde von Mousing et al. (1997 [181]) vorgestellt. Die Diagnose von Salmonella enterica*-Infektionen erfolgt mittels Fleischsaft-ELISA zum Zeitpunkt der Schlachtung und zielt auf eine Einteilung der Herkunftsbetriebe in drei verschiedene Risikoklassen ab. In den Risikoklassen gelten verschiedene regulatorische Maßnahmen, die hier nicht weiter ausgeführt werden sollen.*

[8] In der medizinischen Terminologie werden Screening-Tests und Diagnosetests mitunter voneinander unterschieden, obwohl eine klare Grenze zwischen beiden nicht immer gezogen werden kann (Walter und Jadad, 1999 [252]). Im vorliegenden Text werden Screening-Tests als Diagnosetests betrachtet, die für eine möglichst vollständige Untersuchung einer (Sub-) Population verwendet werden.

Bei Beispiel 1.2 handelt es sich um eine Aggregattestung, weil auf Grund der Testergebnisse der Erzeugerbetrieb gemaßregelt wird. Die Bedeutung der Salmonellose-Bekämpfung ist auf dem Hintergrund der Zoonoseproblematik zu sehen.

In Anhang B.5 wird ein Überwachungssystem für die Trichinellose beim Schwein vorgestellt, welches unter Verzicht auf die allgemeine Trichinenschau in erster Linie den gesundheitlichen Verbraucherschutz sichern soll. Die wesentlichen Elemente dieses Überwachungsschemas sind Zertifikation von "Trichinella-freien" Betrieben und Gebieten und eine serologische Testung von Masttieren, die aus nicht-zertifizierten Betrieben zugekauft werden. Auf Grund der sehr geringen Prävalenz der Trichinellose beim Schlachtschwein ist allerdings die Aussagekraft positiver Testresultate gering. Dieser Sachverhalt kann durch die Berechnung des prädiktiven Werts eines positiven Testbefunds formal zum Ausdruck gebracht werden (Abschnitt 3.3.1 und S. 202). Eine quantitative Entscheidungsanalyse unter Berücksichtigung der Testeigenschaften, der Prävalenz und der ökonomischen Konsequenzen von falsch positiven Diagnosen wurde beschrieben (Greiner und Baumann, 1998 [87]; Anhang B.5).

1.2.3 Prävalenzstudie

Studien zur Schätzung einer Prävalenz zielen darauf ab, mit einem möglichst geringen Untersuchungsaufwand zu einer ausreichend genauen und unverzerrten Quantifizierung der Verbreitung eines Merkmals (z.B. Infektion, Krankheit, genetischer Defekt, Risikofaktor) in einer definierten Zielpopulation zu gelangen. Ein typisches Studiendesign für Prävalenzschätzungen ist die Querschnittsstudie (Abschnitt 3.3.2). Bei der Verwendung von indirekten Diagnoseverfahren, wie beispielsweise einem serologischen Test, wird häufig die serologische oder *apparente* Prävalenz als eine grobe Schätzung der tatsächlichen Prävalenz angegeben, obwohl bekannt ist, dass das diagnostische Verfahren mit Fehlern behaftet ist (s. hierzu Abschnitt 6.2; Greiner und Gardner, 2000 [98]).

Beispiel 1.3 (Schätzung der Paratuberkuloseprävalenz beim Rind)
Die Prävalenz der Paratuberkulose beim Rind auf Einzeltier- und Herdenbasis in Missouri wurde von Thorne und Hardin (1997 [240]) mittels Absorptions-ELISA untersucht. In einem Teil der Untersuchungen wurden Serumproben, die zum Zweck der amtlichen Brucellose-Testung gezogen wurden als Auswahlrahmen verwendet und ein Stichprobenumfang für Tiere innerhalb der Betriebe so gewählt, dass in jedem positiven Betrieb mit einer Wahrscheinlichkeit von 95% mindestens ein infiziertes Tier nachgewiesen werden konnte. Hierbei wurde die Annahme getroffen, dass in einem positiven Betrieb eine Prävalenz von mindestens 10% vorlag. Die Prävalenzschätzungen wurden für Missklassifikationen korrigiert (Abschnitt 6.2). Die Intracluster-Korrelation wurde

geschätzt und bei der Berechnung eines Vertrauensbereichs berücksichtigt (Abschnitt 3.3.2 und A.2).

Wie später ausgeführt werden wird (Abschnitt 2.2.3), ist die Diagnostik der Paratuberkulose dadurch erschwert, dass bei latenten Ausscheidern Serumantikörper häufig nicht nachgewiesen werden können. Hierdurch besteht die Gefahr einer drastischen Unterschätzung der Infektionsprävalenz in einer latent infizierten Herde.

1.2.4 Epidemiologische Untersuchung von Risikofaktoren

Das Ziel epidemiologischer Studien ist es, den Zusammenhang zwischen der Exposition mit einem Risikofaktor und dem Vorliegen einer Infektion, Krankheit oder einer Leistungsminderung (allgemein als *Zielvariable* bezeichnet) zu quantifizieren. Die Prävalenz und Inzidenz der Risikofaktoren und Zielvariablen in der Population sind die wesentlichen Rahmenbedingungen für die Entscheidung zwischen verschiedenen Studientypen, wie Kohortenstudie, Fall-Kontroll-Studie und Querschnittsstudie (Kreienbrock und Schach, 2000 [143], Zessin und Greiner, 2000 [262]). Die Diagnostik sowohl des Risikofaktors als auch der Zielvariable beruht häufig auf diagnostischen Tests.

Beispiel 1.4 (Risikofaktoren für Lenti-Virus-Seropositivität)
Keen et al. (1997 [139]) untersuchten potenzielle Risikofaktoren für Lenti-Virus-Seropositivität bei weiblichen Zuchtschafen in Nebraska mittels Antikörper-ELISA. Ovine Lentiviren sind die Erreger der Maedi-Visna beim Schaf. Unter Verwendung eines logistischen Regressionsmodells wurden die räumliche Abtrennung der Geburt und Aufzucht sowie ein höheres Absatz- und Lebensalter als Risikofaktoren für die Seropositivität ermittelt.

Bei der zitierten Studie wird der serologische Status als Zielvariable verwendet und muss daher als eine *Surrogatvariable* für den tatsächlichen, nicht beobachteten Infektionsstatus aufgefasst werden, dem wohl das eigentliche Interesse der Studie gilt. Sofern die diagnostischen Testeigenschaften gänzlich unbekannt sind, kann für Testfehler nicht adjustiert werden und es muss offen bleiben, ob die identifizierten Risikofaktoren nur zur Erklärung des serologischen oder auch des tatsächlichen Infektionsstatus beitragen. Im vorliegenden Beispiel könnte sogar vermutet werden, dass der Alterseffekt teilweise auf die Akkumulation von Antikörpern im fortgeschrittenden Alter zurückzuführen ist. Diagnosefehler in epidemiologischen Studien können “echte” Risikofaktoren maskieren oder zu Fehlschlüssen über vermeintliche Risikofaktoren führen (Abschnitt 6.3).

1.2.5 Quantitative Risiko-Analyse

Auf der Grundlage internationaler Handelsabkommen besteht die Möglichkeit, die Einfuhr von lebenden Tieren oder von Erzeugnissen tierischer Produktion zu unterbinden, wenn mit der Einfuhr ein wissenschaftlich gesichertes Risiko für die Einschleppung einer Tierseuche oder die Gefährdung der menschlichen Gesundheit verbunden ist (Artikel 5 des SPS-Abkommens, WTO, 1996 [259]). In diesem Zusammenhang kommt den Diagnosetests eine besondere Bedeutung zu. Zunächst muss das "Grundrisiko" der Tierseuche im exportierenden und importierenden Land bestimmt werden (Prävalenzstudie). Die Genehmigung für Versendung oder Einfuhr von lebenden Tieren ist dann meist von deren serodiagnostischem Status abhängig (Aggregattestung).

Beispiel 1.5 (Import von Rinderembryonen aus Südamerika)
Sudmoller und Wrathall (1997 [237]) führten eine quantitative Risiko-Analyse für den internationalen Handel mit Rinderembryonen aus Südamerika durch, motiviert durch die Verbreitung der Maul- und Klauenseuche, Blauzungenkrankheit und der Vesikulären Stomatitis bei Rindern im Herkunftsgebiet. Die Autoren identifizierten und quantifizierten die wesentlichen Einflussfaktoren und Gefahrenpunkte eines hypothetischen Risikoszenarios. Das Einschleppungsrisiko für die drei genannten Seuchen konnte durch eine stochastische Simulation modelliert werden. Hierbei wurden die Einflussfaktoren und Gefahrenpunkte durch Zufallszahlen repräsentiert, deren statistische Verteilungen mit Hilfe von publizierten Daten, eigenen Überlegungen und Expertenwissen definiert wurden.

In Abschnitt 4.2 wird darauf hingewiesen, dass die diagnostischen Untersuchungen, die solchen Risikoszenarien zu Grunde liegen, als multiple Testverfahren, kombiniert mit einem Aggregattest, aufzufassen sind. Die wirtschaftlichen Konsequenzen für Testfehler sind beachtlich. Bei einer falsch positiven Testung kommt es durch die Zurückweisung gesunder Tiere für den Exporteur zu einem Verlust an Absatzmöglichkeiten. Weitaus schwerwiegender können allerdings die Folgen falsch negativer Testresultate sein, wenn es hierdurch tatsächlich zu einer Verbreitung einer Tierseuche kommt.

http://www.springer.com/978-3-540-00401-1

Serodiagnostische Tests
Evaluierung und Interpretation in der Veterinärmedizin und anderen Fachgebieten
Greiner, M.
2003, XVIII, 242 S., Softcover
ISBN: 978-3-540-00401-1